Chapter 4

Measures of Dispersion:

Standard deviation

Coefficient of Variation

Prepared by, ***Jannatul Ferdous,*** *Assistant Professor, Dept. of CSE, Metropolitan University, Sylhet*

Measures of Dispersion

The value given by a measure of central tendency is considered as the representative of the whole data and can describe only one of the important characteristics of a series. It does not give the spread or range over which the data are scattered. Measures of dispersion used to indicate this spread and the manner in which data are scattered. Consider the following example:

Profit (in lac Tk.) of two companies for first six months in the year 2010 is given below:

From the above information, it has been seen that the companies have an equal average of profit which may indicates that both the companies are same. It fails to indicate the differences.

However, if we see the observations of the companies it has easily seen that the profit of company A is more varying or fluctuating comparing to its mean profit. On the other hand, the profit of company B is more stable or less varying compare to its mean profit. Therefore, company B is more stable than the company A which is possible to conclude by using dispersion not by the average.

Therefore, to analyze statistical data dispersion or variation is more important.

Company	Jan	Feb	March	April	May	June	Average
A	16	5	12	0	9	30	12
B	10	11	13	9	14	15	12

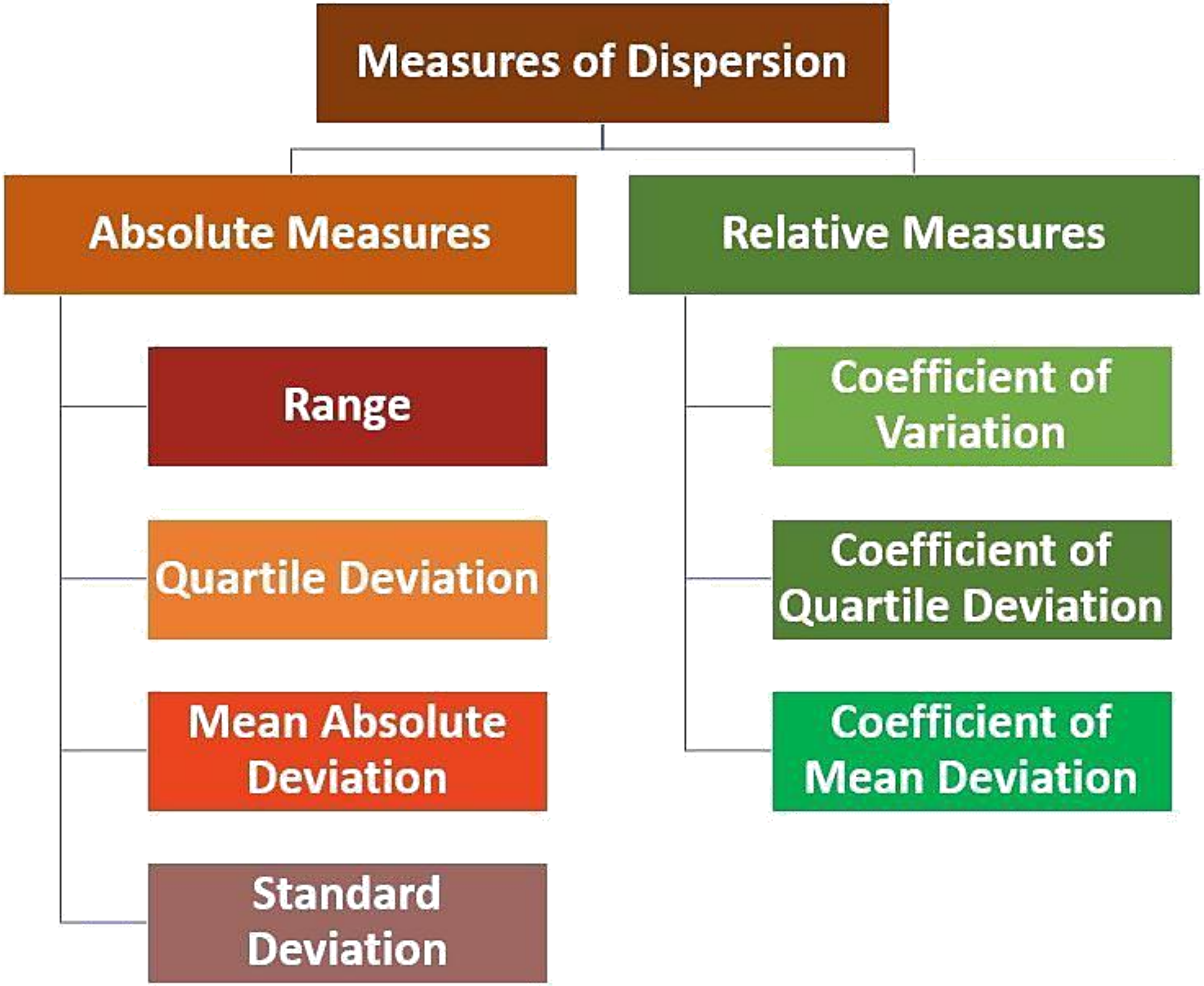
There are many examples of variance and standard deviation. Variance, standard deviation, range, inter-quartile range are all measures of spread of data. That tells you how far away data is from the middle/center of the data.

Example 1: Stock market or other investment returns. The stock market has return on average 7% per year. This does not mean that every year you get a 7% return, some years are more and some years are less. This variability (called volatility in stock terms) is an example of variance and standard deviation.

Example 2: A manufacturing machine fills up water bottles with 8oz. Now the machine does not fill up each water bottle with exactly 8oz, there is a little variation due environmental factors (weather, placement of bottle, etc.). This variation is example of variance or standard deviation.

Example 3: You drive to work every day and take the same route. There is both variation in the time is takes you to get to work (traffic, stop light timing etc.) and variation in the amount of gas you use (same factors of traffic, stop light time affect this). All of this variability can be measured with variance and standard deviation.
Basically, anywhere you see movement away from the center of something it can be measured with variance or standard deviation (for numerical things anyway).

Method of Measures of Dispersion



Variance (σ^2) is a measurement of the spread between numbers in a data set. That is, it measures how far each number in the set is from the mean and therefore from every other number in the set. It is defined as,

Variance, $\sigma^2 = \frac{\sum_{i=n}^{n}(x_i - \bar{X})^2}{n}$

where: x_i = the i^{th} data point

$\bar{x}$ = the mean of all data points

n = the number of data points

Standard deviation is a statistic that measures the dispersion of a dataset relative to its mean and is calculated as the square root of the variance (σ^2). If the data points are further from the mean, there is a higher deviation within the data set; thus, the more spread out the data, the higher the standard deviation.

Standard deviation (SD), $\sigma = \sqrt{\frac{\sum_{i=n}^{n}(x_i - \bar{X})^2}{n}}$

For population (ungrouped data)

Variance, $\sigma^2 = \frac{\sum_{i=n}^{n}(x_i - \bar{X})^2}{n}$

Standard deviation (SD), $\sigma = \sqrt{\frac{\sum_{i=n}^{n}(x_i - \bar{X})^2}{n}}$

For sample (ungrouped data)

Variance, $S^2 = \frac{\sum_{i=n}^{n}(x_i - \bar{X})^2}{n-1}$

Standard deviation (SD), $S = \sqrt{\frac{\sum_{i=n}^{n}(x_i - \bar{X})^2}{n-1}}$

For population (grouped data)

Variance, $\sigma^2 = \frac{\sum_{i=n}^{n} f_i(x_i - \bar{X})^2}{n}$

Standard deviation (SD), $\sigma = \sqrt{\frac{\sum_{i=n}^{n} f_i(x_i - \bar{X})^2}{n}}$

For sample (grouped data)

Variance, $S^2 = \frac{\sum_{i=n}^{n} f_i(x_i - \bar{X})^2}{n-1}$

Standard deviation (SD), $S = \sqrt{\frac{\sum_{i=n}^{n} f_i(x_i - \bar{X})^2}{n-1}}$

Working formula

For population (ungrouped data)

Standard deviation (SD), $\sigma = \sqrt{\frac{\sum(x_i)^2}{n} - \left(\frac{\sum x_i}{n}\right)^2}$

For sample (ungrouped data)

Standard deviation (SD), $S = \sqrt{\frac{\sum(x_i)^2}{n-1} - \left(\frac{\sum x_i}{n-1}\right)^2}$

For population (grouped data)

Standard deviation (SD), $\sigma = \sqrt{\frac{\sum f_i(x_i)^2}{n} - \left(\frac{\sum f_i x_i}{n}\right)^2}$

For sample (grouped data)

Standard deviation (SD), $S = \sqrt{\frac{\sum f_i(x_i)^2}{n-1} - \left(\frac{\sum f_i x_i}{n-1}\right)^2}$

Example 1: Calculate the standard deviation of the following data

Age (X)	20-25	25-30	30-35	35-40	40-45	45-50	50-55
No. of workers (f_i)	17	11	8	5	4	3	2

Solution:

Age (X)	**No. of workers (f_i)**	**Mid value (X_i)**	$f_i x_i$	$f_i {x_i}^2$
20-25	17	22.5	382.5	8606.25
25-30	11	27.5	302.5	8318.75
30-35	8	32.5	260	8450
35-40	5	37.5	187.5	7031.25
40-45	4	42.5	170	7225
45-50	3	47.5	142.5	6768.75
50-55	2	52.5	105	5512.5
	$\sum f_i =$ **n = 50**		$\sum f_i x_i =$ **1550**	$\sum f_i {x_i}^2 =$ **51912.5**

Standard deviation (SD), $\sigma = \sqrt{\frac{\sum f_i (x_i)^2}{n} - \left(\frac{\sum f_i x_i}{n}\right)^2} = \sqrt{\frac{51912.5}{50} - \left(\frac{1550}{50}\right)^2} = \sqrt{1038.25 - 961}$

$$= \sqrt{1038.25 - 961} = \sqrt{77.25} = 8.789 = 8.79$$

Coefficient of Variation

In statistics, the coefficient of variation, also known as relative standard deviation, is a standardized measure of dispersion of a frequency distribution. It is often expressed as a percentage, and is defined as the ratio of the standard deviation to the mean.

$$C.V. = \frac{standard\ deviation}{arithmatic\ mean} = \frac{\sigma}{\mu} \times 100$$

σ = standard deviation
μ = population mean
, for population

$$= \frac{S}{\bar{x}} \times 100$$

S = standard deviation
$\bar{x}$ = sample mean
, for sample

Example 2: Coefficient of variation of two distributions are 60% & 70% and their standard deviations are 21 & 16, respectively. What are their arithmetic means?

Solution:

For first distribution

Coefficient of variation(CV) = 60

Standard deviation = 21

We know that

$$CV = \frac{Standard\ deviation}{Mean} \times 100$$

$$60 = \frac{21}{Mean} \times 100$$

$$Mean = \frac{21}{60} \times 100$$

Mean = 35

For second distribution

Coefficient of variation(CV) = 70

Standard deviation = 16

We know that

$$CV = \frac{Standard\ deviation}{Mean} \times 100$$

$$70 = \frac{16}{Mean} \times 100$$

$$Mean = \frac{16}{70} \times 100$$

Mean = 22.85

The lower the value of the coefficient of variation, the more precise the estimate.

Example 3: Average price of **Stock A** was $50 with a standard deviation of $5 and average price of **Stock B** was $100 with a standard deviation of $5. Calculate the coefficient of variation of two Stocks and comment on its variability.

Solution:

Stock A:

- Average price last year = $50
- Standard deviation = $5

$$CV_A = \left(\frac{S}{\overline{X}}\right) \cdot 100\% = \frac{\$5}{\$50} \cdot 100\% = 10\%$$

Stock B:

- Average price last year = $100
- Standard deviation = $5

$$CV_B = \left(\frac{S}{\overline{X}}\right) \cdot 100\% = \frac{\$5}{\$100} \cdot 100\% = 5\%$$

Both stocks have the same standard deviation, but stock B is less variable relative to its price